

The Shape of the Search Tree for the Maximum Clique Problem, and the Implications for Parallel Branch and Bound

Ciaran McCreesh* Patrick Prosser
 c.mccreesh.1@research.gla.ac.uk patrick.prosser@glasgow.ac.uk

September 5, 2014

Abstract

Finding a maximum clique in a given graph is one of the fundamental NP-hard problems. We compare two multi-core thread-parallel adaptations of a state-of-the-art branch and bound algorithm for the maximum clique problem, and provide a novel explanation as to why they are successful. We show that load balance is sometimes a problem, but that the interaction of parallel search order and the most likely location of solutions within the search space is often the dominating consideration. We use this explanation to propose a new low-overhead, scalable work splitting mechanism. Our approach uses explicit early diversity to avoid strong commitment to the weakest heuristic advice, and late resplitting for balance. More generally, we argue that for branch and bound, parallel algorithm design should not be performed independently of the underlying sequential algorithm.

1 Introduction

When parallelising an algorithm, we seek to divide work into a suitable number of equally sized, independent subproblems which may be evaluated simultaneously. Sequential branch and bound algorithms do not lend themselves to this approach, and parallel branch and bound algorithms must cope with extremely irregular subproblem sizes, speculative evaluation and unpredictable communication. Nonetheless, parallel branch and bound can be beneficial in practice. Here we look at some of the choices available when parallelising a branch and bound algorithm. We show that different work splitting strategies can have a substantial effect upon performance, and that a careful study of the underlying sequential algorithm can help direct our design.

We look at parallel algorithms for the maximum clique problem, and make use of the extra programming flexibility offered by multi-core systems to take measurements “inside the search” to demonstrate *why* some implementation choices give better or worse performance on some problems. We make three claims, and justify them experimentally. Firstly, we show that static work splitting can lead to balance problems. Secondly, we show that different search orders and work splitting mechanisms often have a large effect upon speedups, and that this is due to changes in “the amount of work done”, rather than just imbalances and overheads. Thirdly, we show that an understanding of how ordering heuristics behave can explain why a simple fixed-depth work splitting mechanism usually does so well, despite occasional balance problems. Using these results, we discuss how we may improve work balance whilst retaining the search-order benefits of a simpler work splitting mechanism. We present a novel work splitting mechanism which explicitly diversifies at the top of search, to offset poor early heuristic advice, and that resplits work later to improve balance. We show that, as our explanations predict, this gives better and more consistent speedups than other mechanisms.

1.1 The Maximum Clique Problem

A clique in a graph is a set of vertices, each of which is adjacent to every other vertex in this set. The maximum clique problem is to find the largest such set in a given graph (see Figure 1 on the following page). This is one of Garey and Johnson’s fundamental NP-hard problems (1990); applications include

*This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/K503058/1]

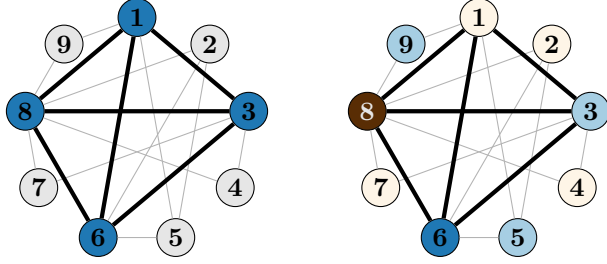


Figure 1: On the left, a graph, with its unique maximum clique $\{1, 3, 6, 8\}$ of size 4 highlighted. On the right, the same graph coloured greedily using four colours: vertices $\{1, 2, 4, 7\}$ are coloured pale cream, $\{3, 5, 9\}$ are light blue, $\{6\}$ is medium blue, and $\{8\}$ is dark chocolate. From this we may infer that there is no clique of size larger than 4. We also know that any clique of size 4 must contain vertices 6 and 8, plus one light blue vertex and one pale cream vertex. This kind of reasoning forms the basis of the sequential algorithms.

bioinformatics (Eblen et al., 2011), biochemistry (Butenko and Wilhelm, 2006), computer vision, electrical engineering, communications (Bomze et al., 1999) and controlling flying robots (Regula and Lantos, 2013). Here we primarily consider dense, computationally challenging graphs (Cheeseman et al., 1991).

The starting point for state-of-the-art exact maximum clique solvers is a series of branch and bound algorithms by Tomita et al. (2003; 2007; 2010). San Segundo et al. (2011; 2013) observed that a bitset encoding could substantially speed up the implementation of these algorithms, without affecting the steps taken; a computational study by Prosser (2012) analyses variations of these algorithms in depth. Subsequently, Batsyn, Maslov et al. showed that substantial improvements could be had by priming these algorithms with a good heuristic solution (Batsyn et al., 2014; Maslov et al., 2013). (Other exact approaches have been tried, including constraint programming (Régin, 2003) and SAT (Li et al., 2011); we focus upon dedicated algorithms here.)

Recently the authors implemented a threaded version of one of these algorithms (McCreesh and Prosser, 2013b). Independently, Depolli et al. did the same (Depolli et al., 2013). Both approaches started with a similar sequential algorithm, produced an implementation using C++11 native threads that explored subtrees in parallel, and performed experiments on similar multi-core hardware. At a glance, both sets of results are similar: good speedups are achievable in practice (but the speedups achieved vary substantially between problem instances), and superlinear speedups sometimes happen—in other words, both papers show that adding parallel tree-search to a strong maximum clique algorithm is worth doing. A close inspection of both sets of results shows that for some particular problem instances, there are substantial differences: for example, the authors obtained roughly linear speedups for the DIMACS graph “MANN_a45”, whereas Depolli et al. achieved a speedup of below 4 regardless of the number of processor cores available. We will explain why this happened, and how this may be addressed without compromising other results.

We begin with a brief recap of sequential maximum clique algorithms. We show how branch and bound algorithms may be viewed as a tree search process, and how we can treat behaviour “inside search” as different kinds of subtrees. We use this to explain the fundamental concepts underlying parallel branch and bound, and discuss the potential for speedups.

In section 3 we look in more detail at the effects of design choices in practice. We show that the reason for Depolli et al.’s speedup limit of 4 for “MANN_a45” is due to poor load balancing. We then look in more detail at further problems. We will see that that balance does not severely restrict speedups in most cases, and that steps taken to improve the balance will often give much worse performance (and not just due to overheads). Our view of branch and bound as a tree, together with some knowledge of how heuristics behave, let us explain this. We show that diversity (that is, using parallelism to hedge against weak early heuristic choices), not balance, is the primary contributing factor to performance in most of the cases we investigate.

Finally, in section 4 we discuss how to get “the best of both worlds”. Driven by a deeper understanding of the search process, we present a new work splitting technique which explicitly adds diversity early in search, to offset the weakest heuristic advice, followed by a low-overhead resplitting mechanism to even out balance problems later on in search.

Algorithm 1: An outline of a maximum clique algorithm, with many details omitted.

```

expand (Graph  $G$ , Set  $C$ , Set  $P$ , Set  $C_{max}$ )
1 begin
2   Construct a greedy colouring of the subgraph induced by  $P$ 
3   for  $v \in P$  (in reverse colour order) do
4     if  $|C| + \text{number of colours used to colour } P > |C_{max}|$  then
5        $C \leftarrow C \cup \{v\}$  // Add  $v$  to growing clique
6        $P' \leftarrow P \cap N(G, v)$  // Filter non-adjacent vertices
7       if  $P' = \emptyset$  then
8         if  $|C| > |C_{max}|$  then  $C_{max} \leftarrow C$  // Unseat the incumbent
9       else expand( $G, C, P', C_{max}$ ) // Recurse with  $v \in C$ 
10       $C \leftarrow C \setminus \{v\}$  // Now consider  $v$  not in the clique
11       $P \leftarrow P \setminus \{v\}$ 

12 maxClique (Graph  $G$ )  $\rightarrow$  Set
13 begin
14   Permute  $G$  to be in non-increasing degree order
15    $C_{max} \leftarrow \emptyset$ 
16   expand( $G, \emptyset, V(G), C_{max}$ )
17   return  $C_{max}$ 

```

2 Algorithms for the Maximum Clique Problem

Throughout, let $G = (V, E)$ be a graph with vertex set V and edge set E , and let $v \in V$. We may write $V(G)$ for V . The *neighbourhood* of v , denoted $N(G, v)$, is the set of vertices adjacent to v . The size of a maximum clique is denoted ω . A *colouring* is an assignment of vertices to colours, such that adjacent vertices are given different colours (as in Figure 1 on the previous page).

We observe that a clique in G either does not contain v , or contains only v and vertices adjacent to v . This allows us to build cliques by recursively selecting a vertex which could be added to a growing clique, and branching on whether or not to include it. We keep track of the largest clique found so far, which we call the *incumbent*. We use the size of the incumbent as a bound: if we can show that a growing clique cannot possibly unseat the incumbent, we may abandon search and backtrack. We do this by constructing a greedy graph colouring—if a subgraph can be coloured using k colours, it cannot possibly contain a clique of size greater than k (each vertex in a clique must be given a different colour).

The sequential algorithms underlying both the authors’ parallel approach (McCreesh and Prosser, 2013b) and that by Depolli et al. (2013) are based upon these principles. We give a brief outline of these algorithms, and refer the reader to the papers by Tomita et al. (2003; 2007; 2010) and San Segundo et al. (2011; 2013) for detailed analysis. We note that the vertex ordering at the top of search differs slightly between our two approaches, so runtimes and node counts are not directly comparable. Prosser’s computational study (2012) discusses these choices in more detail; for our purposes, it suffices to know that initial vertex ordering can make a substantial difference in the size of the search space, but that there is no “best” ordering.

2.1 Sequential Branch and Bound

In Algorithm 1 we give the underlying sequential procedure. The variable C is a growing clique, and P contains undecided vertices which could potentially be added to C . The best solution found so far is stored in C_{max} . The important part, for this discussion, is that we have a **for** loop, and that the **expand** function calls itself recursively inside this loop. Some of these recursive calls are then avoided, by using a greedy graph colouring to prove that the current portion of the search space could not possibly lead to a better solution than the one in C_{max} being found.

We may view the recursive calls made by a branch and bound algorithm as forming a tree, as in Figure 2 on the next page. Here, each node in the tree represents a recursive call; triangles represent

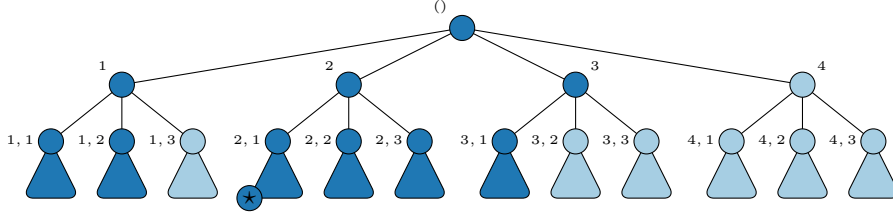


Figure 2: A possible search space of a branch and bound algorithm, viewed as a tree—we assume a depth first search, iterating over children from left to right. Each node represents a recursive call; triangles represent large subtrees. The optimal solution is marked \star , at location $(2, 1 \dots)$. Nodes which need not be explored if this optimal solution has already been found, which are said to be eliminable, are shaded in light blue; nodes which must always be explored to prove optimality are shaded in medium blue. Here, the light blue subtree $(1, 3)$ is avoidable—it would be explored in the sequential run despite being eliminable.

large subtrees. We mark by \star the location of an optimal solution. Nodes shown in medium blue are those which cannot be eliminated by the bound, regardless of the strength of the incumbent—we say such nodes are *ineliminable*. Nodes in light blue are those which could be eliminated by the bound, if \star has been found—such nodes are *eliminable*. The tree is traversed in a manner similar to a depth-first search, exploring subtrees from left to right. Note that in a sequential run, the leftmost light blue subtree will *not* be eliminated by the bound, since the search will not yet have found \star . In other words, not all eliminable nodes are necessarily eliminated. We call eliminable nodes that are not eliminated in the sequential run *avoidable*.

In maximum clique terms, \star is the location of a maximum clique, whose size is denoted ω . The medium blue nodes are then the nodes which must be explored to prove that there is no clique of size $\omega + 1$ in the graph. Light blue nodes are those that would be eliminated by the bound, if the algorithm were to be initialised with an incumbent size of ω rather than 0. For now we assume there is a unique optimum, and that the optimum must be known before any subtree may be eliminated by the bound. We will see later that this assumption is slightly too simple in some cases.

We may label nodes with lists of numbers as follows: the root node is labelled with the empty list, $()$. We label child of the root node from left to right as (1) , (2) , (3) , etc. The children of (1) are labelled $(1, 1)$, then $(1, 2)$, $(1, 3)$, and so on. Thus, in our figure, the optimal solution \star is found at the location $(2, 1 \dots)$, which corresponds to taking the second branch at the top of search followed by the first branch thereafter.

2.2 Parallel Branch and Bound, and the Potential for Speedup

Usually when we wish to parallelise an algorithm, we look for independent units of work which may be evaluated independently. With branch and bound we cannot take this simple approach, and we must resort to speculatively executing non-independent subproblems to be able to exploit parallel hardware. We view branch and bound as a tree search, ignore left-to-right dependencies, and explore different subtrees in parallel. There is a single incumbent which is shared between every worker. This technique in general is well known (see, for example, discussions by Bader et al. (2005) and Crainic et al. (2006)). For the maximum clique problem, it was first attempted by Pardalos et al. (1998), where message passing was used on a cluster of four machines, and later, by the authors (McCreesh and Prosser, 2013b) and Depolli et al. (2013) using threads. Another recent parallel maximum clique algorithm by Rossi et al. (2013) does the same, starting with an algorithm designed for large sparse graphs; again, superlinear speedups were observed. Parallel tree-search was also used by Debroni et al. (2011) to close the Keller maximum clique problem, using an algorithm which exploited special properties of these graphs.

Other approaches have been considered. Xiang et al. (2013) used MapReduce rather than conventional parallel branch and bound. Their emphasis is upon the scalability of their solution: they demonstrate linear speedups (over a parallel implementation, not a sequential one) on three graphs from one DIMACS family, and on one further graph. We will see that other families of graphs have properties which can interfere with this approach. Bergman et al. (2014) used parallel decision diagrams, and showed excellent scalability—however, their sequential runtimes were typically much worse than state of the art algorithms.

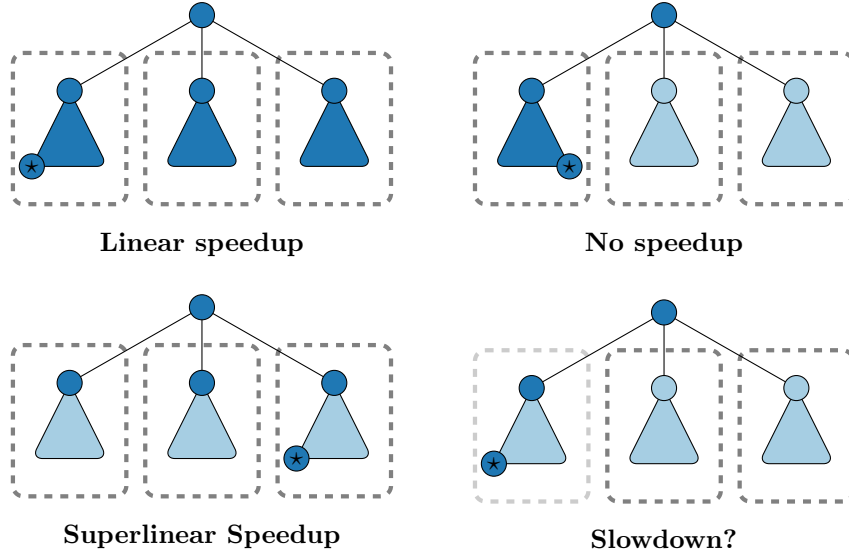


Figure 3: Possibilities for speedup in parallel branch and bound. Here we show the search space being divided between three workers. If there are no eliminable nodes, we are dividing up a fixed amount of work, and may hope for a linear speedup. If our workers spend their time exploring avoidable nodes, as in the top right example, we would get no speedup at all. Conversely, a worker may find an optimal solution much more quickly than in the sequential run. This may lead to avoidable nodes being eliminated in the parallel run, giving a superlinear speedup. Finally, the bottom right example shows the perils of exploring the tree “in a different order” to the sequential run: if only two workers are available, and exploring the leftmost subtree is deferred, we would explore many more nodes in total than in a sequential run.

With the parallel tree-search approach, we are not dividing a fixed amount of work between processor cores. Thus we should not always expect to gain a speedup approaching n from n cores (such a speedup is said to be *linear*). Instead, the speedup obtained will depend not just upon the algorithm used, but also the nature of the input. There are various possible outcomes, which we illustrate in Figure 3:

- The search space consists entirely of nodes that cannot be eliminated—discovering the optimal solution does not provide any benefit to proving optimality. In this case, we *are* dividing a fixed amount of work up between workers, and may hope for a linear speedup.
- The search space contains many eliminable nodes that are eliminated on the sequential run, but are not eliminated in the parallel run. Our additional workers end up contributing nothing to the solution, so we get no speedup at all.
- The search space contains many avoidable nodes (eliminable nodes that are not eliminated on the sequential run) that are eliminated in parallel due to one of the additional workers finding an optimal solution quickly. Here we may get a superlinear speedup (Lai and Sahni, 1984; Trienekens, 1990).
- We could start by allocating all our workers to explore portions of the search space that would be eliminated in the sequential run. This could lead to a slowdown. Intuitively, one might think this is due to the tree being explored “in a different order” in parallel. This possibility is indeed avoidable—we must ensure that incumbents are discovered at least as quickly in parallel as they are in the sequential run, and that newly discovered incumbents are shared between threads. We refer to works by Li and Wah (1986), Trienekens (1990) and de Bruin et al. (1995) for details (the injectivity requirements are not relevant for the maximum clique problem, since all cliques of a given size are equivalent for inference purposes).

Both our approach and that by Depolli et al. meet the conditions for avoiding a slowdown (with one technicality: it is possible for colourings of a subproblem to be worse than colourings of a parent

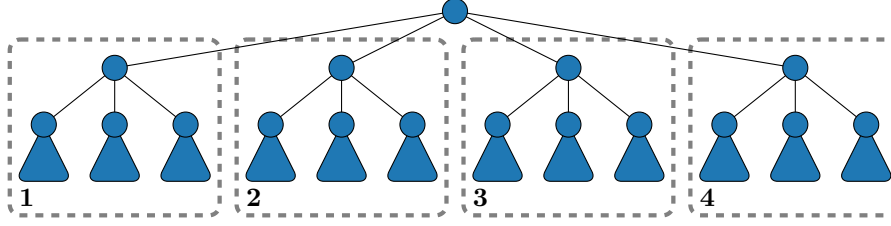


Figure 4: Splitting work at the top: we divide the work into four jobs by splitting the tree at distance 1 from the root. Workers tackle the subproblems in the order shown.

problem—we have been unable to observe this having an effect in practice, although it can easily be worked around (McCreesh and Prosser, 2013a)). The approach by Xiang et al. (2013) does *not* meet these conditions: they do not preserve sequential search order, and they do not share newly discovered incumbents until a subproblem has finished. We will see in the following section why this was not a problem for the three DIMACS graphs that they considered.

Note that these properties may only be categorised after the fact—when the algorithm finds the optimal solution, it does not yet know that there is nothing better. Furthermore, in practice, we would not expect problems to fall neatly into one of these categories—we could both explore some additional eliminable nodes before a solution is found, and avoid some avoidable nodes. The matter is further complicated by the possibility of multiple solutions, and of “quite good” solutions that allow some but not all of the eliminable nodes to be eliminated. Our outcomes only represent the extremes in behaviour that we might observe.

3 Do Details of Parallel Algorithm Design Matter?

Having discussed what could happen in theory, we now show that these concerns are often relevant in practice. But first we briefly discuss our experimental setup and test data.

3.1 Experimental Setup and Data

Both the authors’ previous paper and Depolli et al. considered problem instances from the Second DIMACS Implementation Challenge¹. We also considered random graphs and graphs from BHOSLIB (“Benchmarks with Hidden Optimum Solutions for Graph Problems”)²; Depolli et al. also considered protein product graphs³. We will use all three families here.

We work with a Linux machine with four AMD Opteron 6366 HE processors. Each of these has eight modules, each containing two cores, for a total of 64 cores. (Note that resources are shared between pairs of cores in a module, so it is not the case that we have 64 times as much processing power as is found in a single core used on its own.) The AMD Turbo CORE feature was disabled, to allow us to investigate scalability effects. We emphasise that each core individually is not especially fast, compared to the systems previously used by either the authors or Depolli et al. in their experiments—our goal in this paper is not to solve new instances or to solve existing instances faster, but rather to look at the consequences of algorithm design choices. Speedups are given over a sequential implementation, not a parallel implementation with one thread. Our compiler is GCC 4.9.0, and we use C++11 native threads.

3.2 The Importance of Good Work Splitting

We have assumed so far that we are able to divide work between processors; we now discuss how to do this. A number of approaches are possible. Both the authors and Depolli et al. (2013) started by splitting work “at the top” of search, as in Figure 4. Since the number of vertices in a graph is expected to be much larger than the number of cores available, this produces more jobs than there are workers. We explicitly placed each subtree of distance 1 from the root onto a queue (using an additional thread, and

¹<http://dimacs.rutgers.edu/Challenges/>

²<http://www.nlsde.buaa.edu.cn/~kexu/benchmarks/graph-benchmarks.htm>

³<http://commsys.ijs.si/~matjaz/maxclique/>

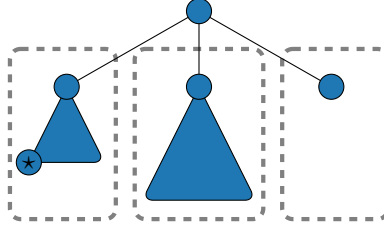


Figure 5: Subtrees can be highly irregularly sized. Here the search space contains no eliminable nodes, so if we divided the work between three processor cores as shown we might expect a speedup approaching three. But if work is not split dynamically, the runtime may be dominated by the cost of exploring the large subtree in the middle.

blocking when the queue contained too many items), and had workers process subtrees from this queue in order. Depolli et al. instead started by processing the root node, and when branching, would process the “take v ” case in one worker, and pass the “do not take v ” case on to a one-item queue to be processed by the next available worker. (We also had a work donation mechanism for further splitting later on in search; we ignore this for now.)

Such an approach assumes that subtrees will be sufficiently numerous and of similar size to allow an even work distribution. We could imagine a situation like the one shown in Figure 5 occurring, where the cost of evaluating one subtree dominates the runtime. Informally, we say that a workload is *balanced* if each thread has roughly the same runtime. We suggest that balance is generally a good thing⁴.

Depolli et al. claim their solution offers “low idle time and good load-balancing”, and for most problem instances this is indeed the case. However, we will now show that a balance problem is the reason they were unable to achieve a speedup of 4 for the DIMACS graph “MANN_a45” regardless of the number of processors used. This contradicts their speculation that the density and large maximum clique size of the graph was to blame for the atypically poor speedup.

Figure 6 on the next page plots speedups obtained by our implementation as the number of worker threads increases (speedups are measured over a good sequential implementation, not a parallel implementation run with one thread). The solid dark blue line shows what happens when splitting work at distance 1, without the work donation mechanism—we see that our speedups appear to be capped at around four⁵.

The first graph in Figure 7 on the following page shows the gap between the runtime of the shortest running and the longest running threads, as the number of threads increases. We would expect the gap between the two lines to be small if a good balance has been achieved; here we see that splitting at distance 1 gives a reasonable balance only for up to four threads.

The dark blue line in the bottom right graph shows runtimes of individual threads, when 32 threads are used. A perfectly balanced work distribution would give a horizontal line. Here we see the longest running thread working for 107s, with the second longest finishing after 79s, and there are only six further threads with runtimes over 10s. The sequential runtime is 438s, and our speedup is capped at $438/107 \approx 4.1$.

One possible workaround is to split the tree further from the root, in the hopes that this will give a better balance. For example, it is trivial to modify our implementation to split at distance 2 from the root, as in Figure 8 on page 9. The solid medium blue lines in Figures 6 and 7 on the following page show that doing so solves the balance problem for “MANN_a45” for up to fifteen threads, beyond which again the runtime is determined by the size of the largest subproblem. Going one step further, and splitting at distance 3 from the root (shown with solid light blue lines) gives a sufficiently balanced distribution to allow effective use of 64 cores. This approach is also taken by Xiang et al. (2013), with a splitting distance chosen at the start of search based upon estimates of the sizes of subtrees.

The principle underlying this approach is that we may solve the irregularity problem by breaking the

⁴Of course, one could cheat and achieve a supposedly perfect balance by having threads that would otherwise be idle perform useless work instead. Thus measures of balance should be used for enlightenment, not for comparisons.

⁵Strictly speaking, it is a coincidence that our limit of “around 4” and Depolli et al.’s figure of “between 3 and 4” both involve the number 4. The two sequential algorithms used different initial vertex orderings, and the implementations give different rates of nodes per second at different depths in the graph, so we should not expect to see exactly the same limit in both cases.

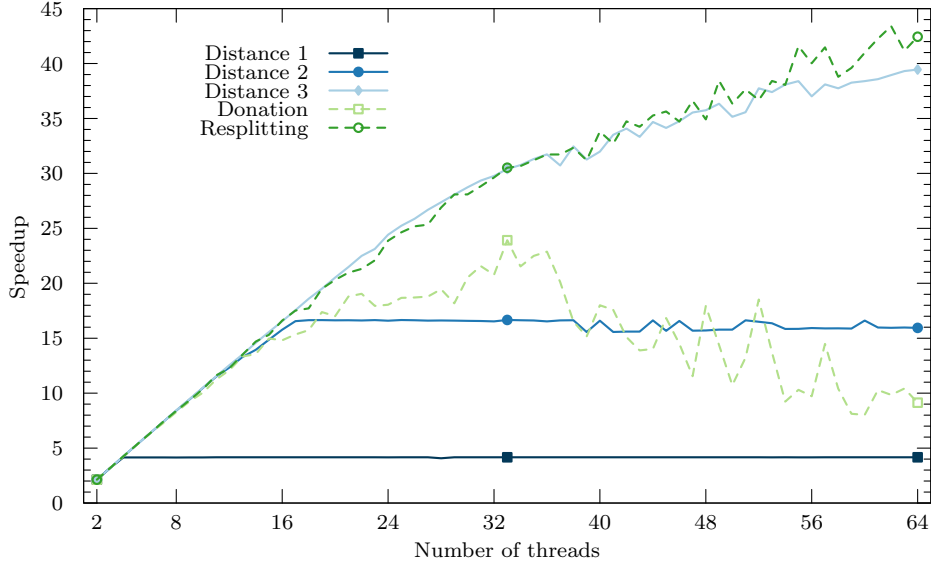


Figure 6: Speedup obtained as the number of threads increases for the DIMACS instance “MANN_a45”. We see that with no work donation, splitting at distance 1 limits the speedup to around 4 regardless of the number of threads used, and splitting at distance 2 limits the speedup to 16. Splitting at distance 3 gives a typical speedup curve. Also shown are results for the authors’ earlier work donation approach (which evidently has scalability limits at this range with this hardware) and a more scalable approach described at the end of this paper. The sequential runtime is 438 s.

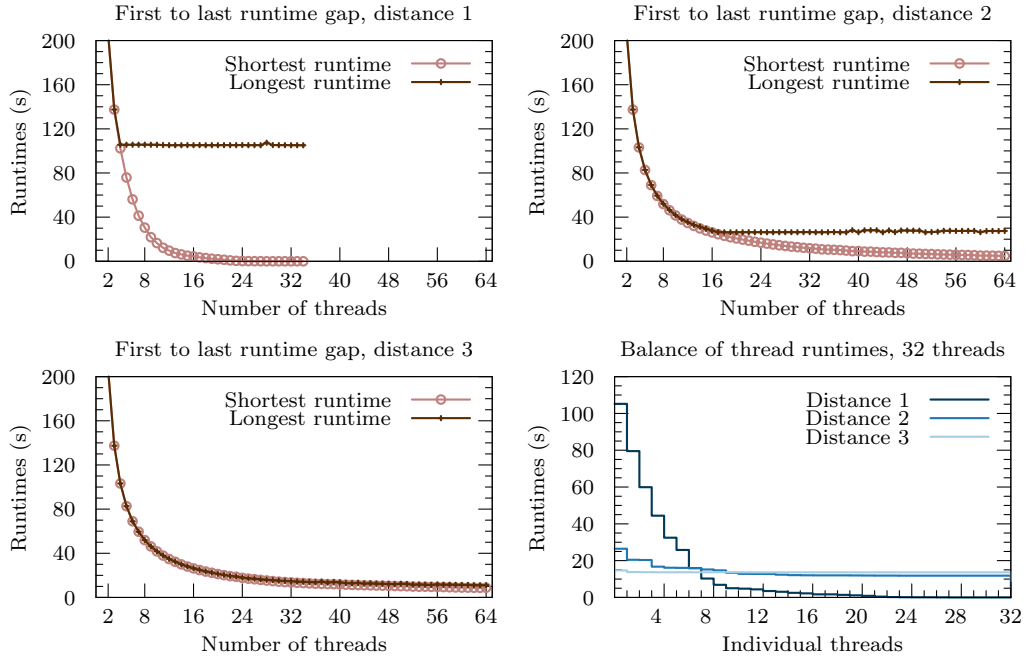


Figure 7: The first three graphs show runtimes of individual worker threads as the number of threads increases for the DIMACS instance “MANN_a45”, with different splitting distances. The top line shows the runtime of the longest thread, and the bottom line the shortest. The fourth graph shows the runtimes of each thread when using 32 threads. We see that as the splitting distance increases, the balance improves.

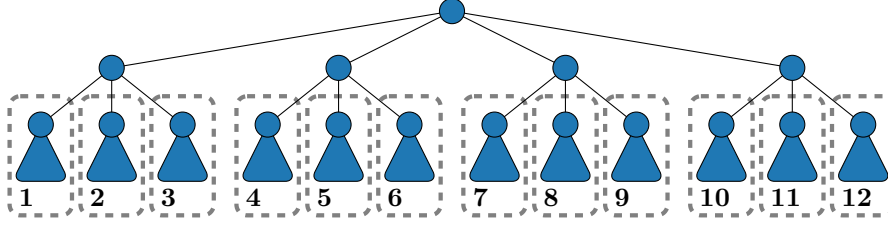


Figure 8: Splitting work at level 2. We might hope that this would lead to a more even work distribution than in Figure 4. But we will see that doing so does not avoid the problems with irregularly sized subtrees, and removes one of the benefits of splitting work at the top.

problem up into many more pieces than we have cores, so that an even balance is obtained automatically. A similar approach in a constraint programming setting has been used by Régim et al. (2013), with very favourable results on a range of standard constraint programming problems (but maximum clique was not considered).

We could also consider splitting work entirely dynamically (Sanders, 1995). If the algorithm is refactored to perform a binary search, this is conceptually simple: before recursing down any “left” branch, the implementation can check whether any thread is out of work, and if one is, then it is given the “right” branch to evaluate. More generally, the splitting could be left to a high level library or language feature—we will investigate this in section 4.2, using Intel’s Cilk Plus. It is important to note that the splitting here is non-deterministic: we will see that this can lead to large variations in runtimes when solving the same problem instance several times.

Unfortunately, for branch and bound algorithms all of these alternatives have extremely undesirable effects. Looking ahead to Figure 9 on page 13 (which uses the same colours for lines), we see the opposite of what is shown in Figure 6 on the preceding page: for other instances, splitting at distance 3 is much worse than splitting at distance 1. This is because deeper splitting does not just affect balance; as a side effect, the search order is also changed. When splitting at distance 1, we explore the subtrees at (1), (2), (3) and so on, in that (parallel) order. But when splitting at distance 2, we would instead start by exploring the subtrees (1, 1), (1, 2), (1, 3) and so on. In a hypothetical situation such as Figure 2 on page 4 where a solution is located at (2, 1, 1...), splitting at distance 1 will lead to a solution being found immediately, but splitting at distance 2 will not. In practice, changing the parallel search order can often have much more significant effects than a potentially improved balance. The remainder of this paper explains this in more detail, and shows how to get close to “the best of both worlds” (these are the light and dark green dashed lines in the graphs).

3.3 Does Parallel Search Order Matter?

Different parallel search orders only matter if we are not dividing up a fixed amount of work—that is, if we are not in the situation shown in the top left of Figure 3 on page 5. If there are not many eliminable nodes, exploring the search tree in any order which does not violate Trienekens’ conditions for avoiding a slowdown is acceptable (Trienekens, 1990). We now show that for many DIMACS instances, there are in fact many *avoidable* search nodes. Thus, even after ensuring that a slowdown cannot happen, parallel search order is an important consideration. One clue to this fact is the recent work by Batsyn, Maslov et al. which “primes” search with a strong incumbent found via iterated local search (Batsyn et al., 2014; Maslov et al., 2013). This technique has yielded substantial improvements for some instances; if there were few avoidable nodes, this would not happen.

In Table 1 on the next page we list the DIMACS instances whose sequential runtime is less than two weeks, then the smaller BHOSLIB instances, and Depolli et al. (2013)’s protein product graph instances. We show the size of a maximum clique, ω , and the number of search nodes (that is, the number of recursive calls made). Next is the number of search nodes that it takes to prove that the graph does not contain a clique of size $\omega + 1$. We find this total by rerunning the algorithm with the size of C_{max} initialised to ω rather than 0. In other words, this is the number of nodes in ineliminable subtrees. We then show the proportion of the search space which consists of avoidable nodes. The final column is discussed below.

Table 1: Properties of the sequential search space for selected problem instances. Shown is the size of a maximum clique ω , then the total number of search nodes. Next is the number of search nodes required to prove there is no clique of size $\omega + 1$, and then the percentage of the search space which is avoidable. Finally, we give the location of the first maximum clique found.

Instance	ω	Total	Prove	Avoid	Location
brock200_1	21	5.25×10^5	3.06×10^5	41.7	22, 4, 10, 6, 1×17
brock200_2	12	3.83×10^3	2.58×10^3	32.6	10, 7, 1×10
brock200_3	15	1.46×10^4	1.45×10^4	0.3	1, 1, 3, 1×12
brock200_4	17	5.87×10^4	3.16×10^4	46.2	36, 20, 5, 1×14
brock400_1	27	1.98×10^8	1.17×10^8	41.0	20, 2, 10, 11, 4, 2, 1×21
brock400_2	29	1.46×10^8	4.84×10^7	66.8	13, 8, 10, 2, 1, 2, 1×23
brock400_3	31	1.20×10^8	1.66×10^7	86.2	14, 10, 1, 4, 2, 1×26
brock400_4	33	5.44×10^7	7.67×10^6	85.9	9, 3, 4, 3, 1×29
brock800_1	23	2.23×10^9	1.76×10^9	21.0	16, 9, 9, 16, 3, 1×18
brock800_2	24	2.24×10^9	1.31×10^9	41.4	23, 65, 1×22
brock800_3	25	2.15×10^9	7.03×10^8	67.3	46, 2, 24, 6, 2, 1×20
brock800_4	26	6.40×10^8	5.09×10^8	20.5	4, 32, 7, 2, 1×22
C125.9	34	5.02×10^4	2.69×10^4	46.4	7, 5, 5, 3, 1, 1, 2, 2, 1×26
C250.9	44	1.08×10^9	9.68×10^8	10.5	4, 32, 4, 9, 3, 5, 2×3, 5, ...
C2000.5	16	1.82×10^{10}	1.82×10^{10}	0.0	1, 21, 19, 26, 5, 1×11
c-fat200-1	12	24	3	87.5	5, 1×11
c-fat200-2	24	24	1	95.8	1×24
c-fat200-5	58	139	27	80.6	28, 1×57
c-fat500-1	14	14	1	92.9	1×14
c-fat500-2	26	26	1	96.2	1×26
c-fat500-5	64	64	1	98.4	1×64
c-fat500-10	126	126	1	99.2	1×126
DSJC500.5	13	1.15×10^6	1.09×10^6	5.2	8, 38, 3, 1×10
DSJC1000.5	15	7.70×10^7	7.67×10^7	0.3	1, 96, 1×13
gen200_p0.9_44	44	1.77×10^6	1.49×10^5	91.6	8, 4, 2, 2, 1×40
gen200_p0.9_55	55	1.70×10^5	2.32×10^3	98.6	4, 3, 1×53
gen400_p0.9_65	65	1.76×10^{11}	7.29×10^9	95.9	11, 7, 3, 2, 1×61
gen400_p0.9_75	75	1.05×10^{11}	1.24×10^8	99.9	17, 8, 3, 4, 1×71
hamming6-2	32	32	1	96.9	1×32
hamming6-4	4	82	81	1.2	1×4
hamming8-2	128	128	1	99.2	1×128
hamming8-4	16	3.65×10^4	3.64×10^4	0.0	1×16
hamming10-2	512	512	1	99.8	1×512
johnson8-2-4	4	24	23	4.2	1×4
johnson8-4-4	14	126	115	8.7	1×14
johnson16-2-4	8	2.56×10^5	2.56×10^5	0.0	1×8
keller4	11	1.37×10^4	1.37×10^4	0.3	1, 4, 1, 5, 1×7
keller5	27	5.07×10^{10}	5.07×10^{10}	0.0	1, 2, 1, 13, 1, 1, 3, 1×20
MANN_a9	16	71	60	15.5	1×16
MANN_a27	126	3.80×10^4	3.78×10^4	0.6	1×6, 3, 1×119
MANN_a45	345	2.85×10^6	2.85×10^6	0.2	1×5, 6, 1×4, 6, 1×334
p_hat300-1	8	1.48×10^3	1.29×10^3	12.9	18, 12, 1×6
p_hat300-2	25	4.26×10^3	2.83×10^3	33.5	7, 18, 4, 4, 1×21
p_hat300-3	36	6.25×10^5	2.46×10^5	60.6	69, 21, 2, 3, 2×3, 1×5, 2, 1×23
p_hat500-1	9	9.78×10^3	9.70×10^3	0.8	3, 18, 1×7
p_hat500-2	36	1.14×10^5	3.96×10^4	65.3	114, 10, 10, 5, 2, 4, 1×10, 2, 1×19
p_hat500-3	50	3.93×10^7	1.56×10^7	60.3	96, 64, 12, 8, 2×3, 1×4, 2, 1×38
p_hat700-1	11	2.66×10^4	1.62×10^4	39.3	259, 4, 2, 1×8
p_hat700-2	44	7.51×10^5	3.79×10^5	49.6	62, 35, 18, 6, 2, 3, 1×38
p_hat700-3	62	2.82×10^8	1.60×10^8	43.3	147, 62, 42, 11, 6, 8, 2, 2, ...
p_hat1000-1	10	1.77×10^5	1.75×10^5	0.8	6, 29, 5, 1×7
p_hat1000-2	46	3.45×10^7	2.18×10^7	36.8	171, 36, 25, 5, 2, 2, 1, 2, ...
p_hat1000-3	68	1.30×10^{11}	3.85×10^{10}	70.5	368, 30, 99, 6, 10, 3, 2, 1, ...

Instance	ω	Total	Prove	Avoid	Location
p_hat1500-1	12	1.18×10^6	9.59×10^5	19.1	226, 75, 4, 1×9
p_hat1500-2	65	2.01×10^9	1.09×10^9	45.5	280, 136, 25, 29, 8, 1, 3, 3, ...
san200.0.7.1	30	1.34×10^4	227	98.3	$2 \times 3, 1 \times 27$
san200.0.7.2	18	464	1	99.8	7, 3, 3, 1×15
san200.0.9.1	70	8.73×10^4	18	100.0	$2, 1 \times 69$
san200.0.9.2	60	2.30×10^5	1.06×10^3	99.5	5, 1, 3, 1×57
san200.0.9.3	44	6.82×10^6	4.19×10^5	93.8	3, 4, 2, 1×41
san400.0.5.1	13	2.45×10^3	1	100.0	20, 5, 3, 1×10
san400.0.7.1	40	1.19×10^5	9.94×10^3	91.7	13, 2, 3, 1×37
san400.0.7.2	30	8.89×10^5	8.32×10^4	90.6	27, 1, 2, 1, 2, 1×25
san400.0.7.3	22	5.21×10^5	6.70×10^4	87.1	3, 4, 2, 1, 1, 2, 2, 1×15
san400.0.9.1	100	4.54×10^6	3.33×10^5	92.7	8, 1×99
san1000	15	1.51×10^5	1	100.0	42, 39, 1×13
sanr200.0.7	18	1.53×10^5	1.26×10^5	17.6	5, 4, 8, 6, 2, 1×13
sanr200.0.9	42	1.49×10^7	1.02×10^7	31.7	8, 15, 2, 4, $2 \times 6, 1 \times 32$
sanr400.0.5	13	3.20×10^5	1.96×10^5	38.8	95, 12, 2, 2, 1×9
sanr400.0.7	21	6.44×10^7	6.40×10^7	0.7	1, 22, 4, 11, 6, 1, 3, 2, 1×13
frb30-15-1	30	2.92×10^8	2.18×10^8	25.5	12, 5, 13, 7, 4, 2, $1 \times 3, 3, \dots$
frb30-15-2	30	5.57×10^8	3.32×10^8	40.4	20, 13, 7, 4, 11, 1, 3, 2, ...
frb30-15-3	30	1.67×10^8	1.00×10^8	40.0	13, 3, 8, 4, 4, 1, 1, 4, ...
frb30-15-4	30	9.91×10^8	4.20×10^8	57.7	55, 10, 12, 4, 2, 1, 2, 1, ...
frb30-15-5	30	2.83×10^8	1.77×10^8	37.3	9, 17, 7, 4, 3, 2, 2, $1 \times 3, \dots$
1KZKA_3KT2A_78	247	247	1	99.6	1×247
1allA_3dbjC_41	346	675	372	44.9	1×346
1f82A_1zb7A_5	500	716	294	58.9	1×500
2FDVC_1PO5A_83	556	1.35×10^3	146	89.2	$2, 1 \times 555$
2UV8I_2J6IA_13107	69	4.26×10^3	461	89.2	5, 3, 1, 1, 3, 2, 1×63
2W00B_3H1TA_10858	143	7.77×10^5	1.23×10^5	84.2	6, 6, 1×141
2W4JA_2A2AD_0	447	890	10	98.9	$1 \times 5, 2, 1 \times 441$
3HRZA_2HR0A_476	563	9.35×10^5	3.23×10^5	65.4	1, 2, 2, 1, 2, $1 \times 4, 2, 1 \times 14, \dots$
3P0KA_3GWLB_0	89	90	3	96.7	1×89
3ZY0D_3ZY1A_110	52	52	1	98.1	1×52

For some DIMACS instances (e.g. “hamming8-4”, “johnson16-2-4”, the “keller” and “MANN” graphs and “sanr400.0.7”), we see that there are very few avoidable nodes, so parallel search order does not matter. We note that for each of these graphs, Batsyn, Maslov et al. failed to obtain a substantial improvement by priming the search (their main sequential algorithm differs slightly, although this does not seem to have much effect here). The DIMACS graphs considered by Xiang et al. are also in this category.

But for more than half of the DIMACS instances, at least a third of the search space is avoidable. For “san400.0.5.1” and “san1000”, the entire search space is avoidable—that is, the bound at the top of search is sufficient to prove optimality, if a clique of size ω has already been found. For the remainder of the “san” family (but not “sanr”), at least 87 % of the search space is avoidable. The “brock”, “gen” and “p_hat” graph families also contain many members with substantial avoidable proportions—some members of these families of graphs also gave extremely large speedups with Batsyn, Maslov et al.’s approach. The BHOSLIB instances are similar: the avoidable proportion is between 25 % and 58 %. Finally, the protein product graphs all have very high avoidable proportions, although in some cases these are coupled with a low total node count. Thus we should expect that parallel search order will often, but not always, be an important factor in determining speedups.

3.4 The Quality of Heuristics, and What This Implies

Harvey and Ginsberg’s limited discrepancy search (1995) is a general tree-search technique which is based upon two principles. Firstly, that when a search fails to find a solution immediately, it is likely that it only made a small number of “wrong turns” (a discrepancy is when search does not follow a heuristic, in an attempt to correct one of these wrong turns). Secondly, they claim that “for many problems the

heuristics are *least* reliable early in the search, before making decisions that reduce the problem to a size for which the heuristics become reliable”. There is a long-standing tradition of focusing upon the first of these two claims, and ignoring the second (Korf, 1996; Walsh, 1997; Prosser and Unsworth, 2011; Moisan et al., 2013). Here we will buck the trend and emphasise the second claim.

One observation that suggests that the second claim might hold for the maximum clique problem is that the heuristics use degree information at the top of search. For many of the DIMACS graphs the degrees of most vertices are fairly similar. This has led to sophisticated tie-breaking mechanisms based upon the sum of neighbouring degrees (Tomita and Kameda, 2007; San Segundo et al., 2014); Prosser’s computational study suggests that sometimes these help, and sometimes they make things worse (Prosser, 2012). We consider the possibility that using parallelism to avoid a strong commitment to the first choice made by a weak early heuristic may be a more fruitful alternative than trying to wring even more information out of the graph at the top of search. Explicitly splitting at distance 1 maximises this effect—by assigning additional workers to go against initial heuristic advice, we minimise early commitments.

We return to Table 1 to justify this. The final column shows the location in the search space of the maximum clique found by the sequential algorithm. If heuristics were strong at the top of search, we would expect the first number to be 1 in most cases, and low in the remainder. Sometimes this is indeed the case—for example, for the graphs “hamming8-4”, taking the first heuristic choice sixteen times in a row leads immediately to an optimal solution. Similarly, the heuristic is perfect for “johnson16-2-4” (taking the first heuristic choice at each level again finds a solution immediately), and for “MANN_a27” and “MANN_a45” it is accurate for the first few levels. For these graphs, the cost is almost entirely in proving optimality. This suggests that having at least one thread preserve search order to avoid a slowdown is worthwhile (although in fact for all of these graphs, there are multiple maximum cliques).

On the other hand, for most graphs the first number in the last column is not a 1, or even a small number. The table shows that Harvey and Ginsberg’s second claim holds for the DIMACS, BHOSLIB and non-trivial protein instances for the maximum clique problem: heuristic information is weak at the top of search, and strong commitment to that information can result in reduced performance. Combined with the high proportion of avoidable nodes in many of these cases, we should expect parallel search order to matter, and for explicit diversity early in search to be beneficial compared to strong early commitment. (Additionally, early diversity remains as good as any other solution in cases where heuristics are strong at all levels, or where there are few avoidable nodes.)

3.5 Selected Results in Depth

We will now look in more detail at the behaviour of different work splitting mechanisms for selected instances, to justify our claims about parallel algorithm design.

In Figure 9 on the following page we show the speedups obtained as the number of threads increases from 2 to 64 for the four members of the DIMACS “brock400” family. We see straight away that we are getting very different behaviour to that shown for “MANN_a45” in Figure 6 on page 8. For “MANN_a45”, splitting at distance 3 was clearly the best option, but here in each case splitting at distance 1 is best, and distance 3 is very poor. This shows that balance is not the deciding factor here.

For each of these graphs we see a sudden jump in speedups from being approximately linear to being visibly superlinear when splitting at distance 1. For “brock400_1”, the jump occurs when going from 19 threads to 20, and the solution is located at (20, 2, 10, 11, 4, 2, 1×21). Thus with distance 1 splitting, the 20th thread very quickly finds an optimal solution and allows much of the search tree to be eliminated. The same behaviour occurs for the other three members of the “brock400” family—in each case, the jump occurs with k threads, where k is the first number in the location of an optimal solution. There is no jump when splitting at distance 2 or distance 3.

For “brock400_3” ($\omega = 31$) we get a second jump as we move to 34 threads. We might think perhaps there is a second optimal solution located at, say, (34, 1×30). In fact it is not this clear-cut: the solutions for all four of the “brock400” graphs are unique, but for “brock400_3” there is a strong (but not optimal) incumbent of size 29 located at (34, 5, 8, 1, 3, 3, 1×23). With 33 threads, the best incumbent found by the equivalent time in the search only has size 24.

Balance does have some effect for “brock400_1”, with over 40 threads—we show this in the fifth graph. The sixth graph shows “brock400_4”, where the balance is even better (the other two graphs are in-between). But in each case, splitting at distance 1 gives a reasonable work distribution and best overall

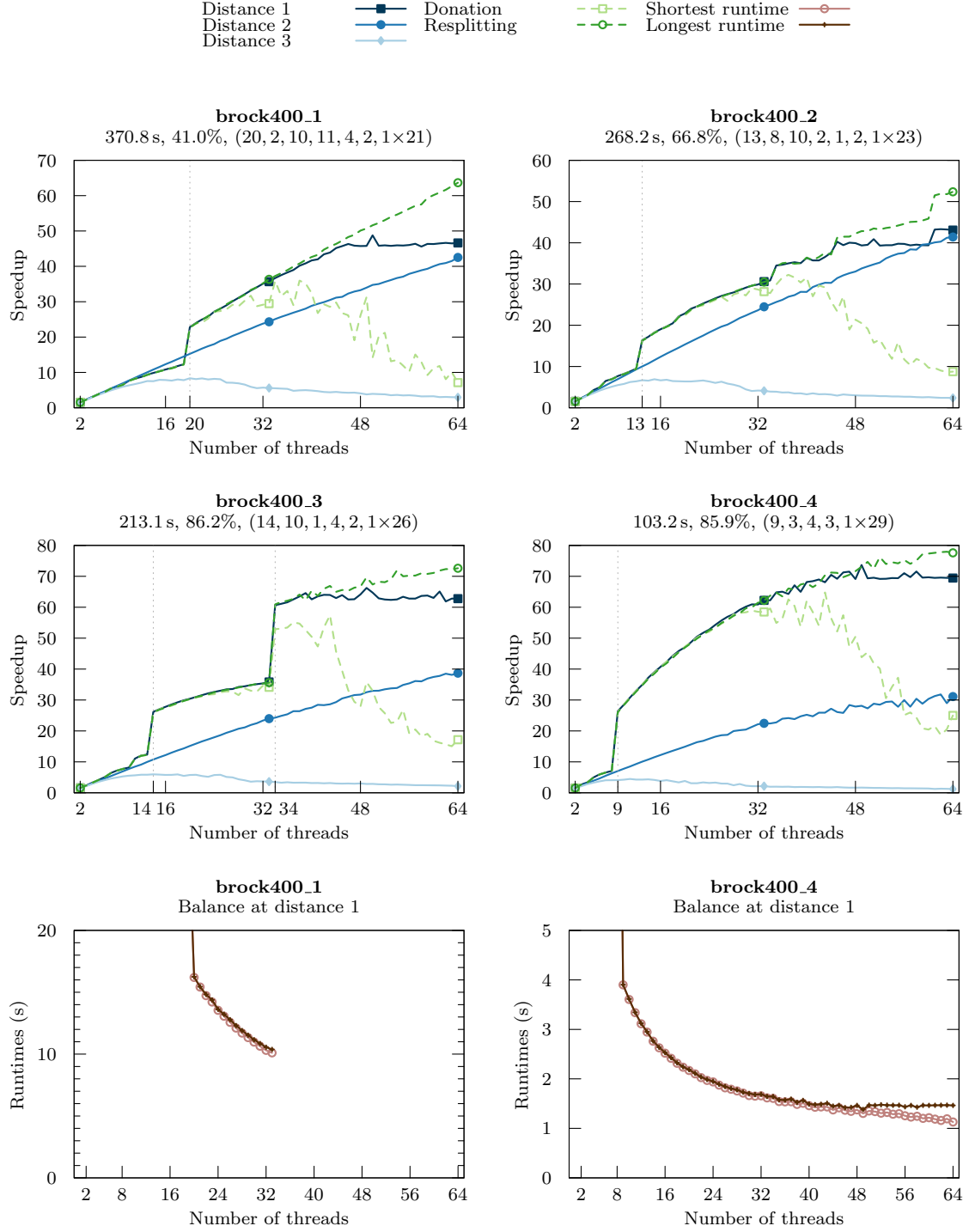


Figure 9: The first four graphs show speedup obtained as the number of threads increases for the “brock400” family of DIMACS instances. Measurements are from a 32 module / 64 core system. The title shows the instance, the sequential runtime, the proportion of avoidable nodes, and the location of the solution found by the sequential algorithm. The dotted vertical lines show where we might expect speedup jumps to occur when splitting at distance 1; the second dotted vertical line in the third graph at position 34 is explained in the text. The bottom two graphs show the balance when splitting at distance 1 with varying number of threads for “brock400.1” (where balance becomes a problem) and “brock400.4” (where balance is better).

performance. We should not find this surprising: Depolli et al. did not encounter hard speedup limits with most instances, so we would not expect balance problems to be common, particularly with smaller numbers of threads.

One further observation is that in each case, for up to around eight threads it does not matter much which mechanism is used—each gives a roughly linear speedup.

But are these results typical? Figure 10 on the following page shows the same information for four further DIMACS graphs. Each shows different interesting behaviour.

For “san400.0.9-1”, where the solution is located at $(8, 1 \times 99)$, there is a very sharp jump as we go from seven to eight threads when splitting at distance 1. Here the eighth thread finds the solution immediately, and this allows over 90 % of the search space to be avoided. We see speedups from nearly 100 to over 300 as we go from 8 to 64 threads. Note that our parallel runtimes go considerably below one second—within this region, our implementation is sensitive to scheduling effects, startup costs and the time spent in the initial sequential part of the problem, and so our speedup lines become rather unstable.

An even stronger superlinear speedup is obtained for “gen400_p0.9-75”, where the speedup from distance 1 splitting is between 800 and 1000 from 32 or more threads. We should expect that this could happen: 99.9 % of the search space is avoidable. But although there is a small jump when going to 17 threads, as we might predict, there are two much larger jumps at 26 and 31 threads. The jump at 17 threads is small because the heuristic is badly wrong at both the top of search, and the second level, so the 17th thread still takes a long time to find a maximum clique. On the other hand, the 26th thread quickly finds an incumbent of size 55 at $(26, 1 \times 3, 2 \times 3, 1, 2, 1 \times 46)$, and a little later the 31st thread finds an incumbent of size 59 at $(31, 1, 3, 4, 1 \times 3, 2, 1 \times 51)$. These are followed by larger incumbents being found by the 17th and 26th threads, and then the 31st thread finds an incumbent of size 72 at $(31, 4, 3, 1 \times 4, 2, 1 \times 64)$. This eliminates enough of the search tree that the maximum clique of size 75 is then found by the 17th thread as we would expect, and search finishes two seconds later. This is an unusually complicated picture, whose behaviour is not captured by simple measurements which assume a single maximum clique and nothing else—we must look more carefully inside the search process to explain what is going on.

This is not the only peculiar behaviour visible in the graph. We can also explain why the speedup decreases slightly as the number of threads goes from 32 to 64. Firstly, we note that there is almost no improvement to the number of nodes required to find an optimal solution when going over 31 threads, so each additional thread could contribute at most a linear extra improvement to the runtime. Secondly, we remind the reader that although our hardware is marketed as having 64 cores, some resources are shared between cores—thus, it does not have 64 times as much processing power when used perfectly in parallel. The result is that although we have more *total* processing power when using all 64 cores, each individual core is penalised somewhat: with 32 threads, stronger incumbents are found slightly sooner than with 64 threads. Since time to find a maximum clique is the most important factor for this instance, the overall speedup is reduced. This is a slightly odd case of what de Bruin et al. (1995) describe as the “[danger of increasing] the processing power of a system by adding a less powerful processing element”; this should also serve as a warning against attempting to gain the benefits of increased diversity by using more threads than there are cores.

For “C250.9” there is no jump. The reasons for this are twofold. Firstly, although the solution is found by taking the fourth heuristic choice at the top of search, the second heuristic decision (where we are not explicitly diversifying) is also poor. Thus, even with distance 1 splitting, we do not find a better incumbent quickly. Secondly, the proportion of avoidable nodes for this instance is relatively low, so finding a stronger incumbent quickly does not provide much benefit. This graph also has a speedup limit from imbalances when splitting at distance 1, which we show in the fifth graph.

Nor is there a jump for “p_hat700-3” ($\omega = 62$). Here splitting at distance 2 beats splitting at distance 1 by a small amount at 16 threads, and a much larger amount by 64 threads. Balance is one factor here, but only from 24 threads onwards. Splitting at distance 1 finds incumbents of sizes up to 53 most quickly, but splitting at distance 2 finds incumbents of sizes 54 and higher in slightly less time. Although 43.3 % of the search space is avoidable, the heuristics in this case are sufficiently poor at the first six levels that we are unable to find an optimal solution quickly regardless of how the work is split.

Overall, we see that balance is sometimes a problem with larger numbers of cores, but that increasing the likelihood of finding an optimal solution quickly is usually far more important. As we predicted, the early diversity offered by splitting at distance 1 is often helpful with this.

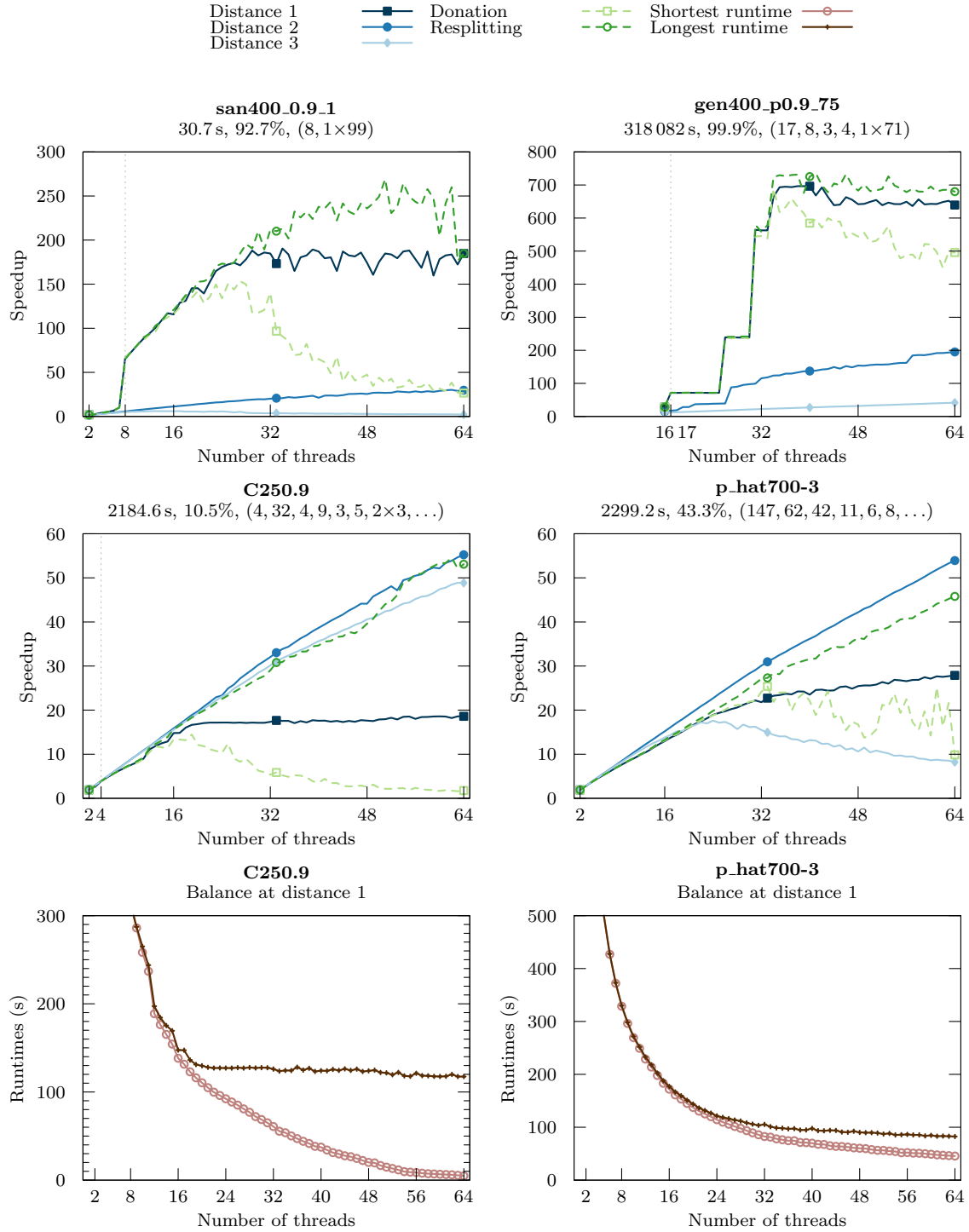


Figure 10: The first four graphs show speedup obtained as the number of threads increases for four further DIMACS instances. Measurements are from a 32 module / 64 core system. The title shows the instance, the sequential runtime, the proportion of avoidable nodes, and the location of the solution found by the sequential algorithm. The dotted vertical lines show where we might expect speedup jumps to occur when splitting at distance 1. For “gen400_p0.9_75”, results using fewer than 16 threads are omitted due to the long sequential runtime. The bottom two graphs show the balance when splitting at distance 1 with varying number of threads for “C250.9” and “p_hat700-3”.

4 Getting the Best of Both Worlds

In view of the previous section, we should look for a work splitting mechanism which gives a good balance, particularly if we are targeting larger numbers of cores. But the results clearly reinforce that splitting at distance 2 or 3 rather than distance 1 has practical consequences beyond the balance of subproblems, and that in most cases parallel search order rather than balance is the dominating factor. We need an approach which gives the best of both worlds.

It is generally known that parallel search order is important if a strong incumbent is not available at the start of search, but selecting a parallel search order explicitly designed to improve our chances of finding a solution quickly has not been given the attention it deserves. For example, Clausen (1997) notes that “most implementations . . . focus on workload distribution methods to increase efficiency of the parallel algorithm”. Clausen discusses balance, and how to assess the efficiency of a parallel algorithm, and observes that “different selection strategies may lead to different search trees”, but does not consider where the solutions are actually likely to be in the search space, and how this can be used to determine a work splitting mechanism. Similarly, in the design of an algorithmic skeleton for parallel branch and bound, Poldner and Kuchen (2008) state that

“the number of problems considered by the parallel skeleton differs enormously over several runs with the same inputs. This number largely depends on the fact whether a subproblem leading to the optimal solution is picked up early or late. Note that the parallel algorithm behaves non-deterministically in the way the search-space tree is explored. In order to get reliable results, we have repeated each run 100 times and computed the average runtimes.”

The search-order issue also occurs in constraint programming. Caniou et al. (2011) note that “solutions may be not uniformly distributed in the search space”, and that this has an affect on parallelism, but they do not consider adapting the search process to improve the chances of finding a solution quickly. The approach by Régim et al. (2013) “relies on the assumption that the resolution time of disjoint subproblems is equivalent to the resolution time of the union of these subproblems”—we have seen that this assumption does not hold here.

When proposing a new work splitting mechanism, Fischetti et al. (2014) state that “since we are interested in measuring the scalability of our method, we considered only instances which are either infeasible or in which we are required to find all feasible solutions (the parallel speedup for finding a first feasible solution can be completely uncorrelated to the number of workers, making the results hard to analyze).” Leroy et al. (2014) do the same in a branch and bound setting: “therefore, we chose to always initialize our B&B by the optimal solution of the instance to be solved. With this initialization, we are sure that the number of explored subproblems is the same in both approaches.” We believe that both of these approaches to measurement and comparison are flawed—parallelism and search order are linked, and should be considered together. In particular, we disagree with Rayward-Smith et al. (1993)’s assertion that speedup is “only a very crude measure of the success of parallelism” and their notion of “pseudo efficiency”. Changes in the amount of work done are inherent to parallelism, and should *not* be disregarded when evaluating a parallelism strategy.

We also desire a stronger notion of reliability than Poldner and Kuchen. For non-trivial problem instances, when splitting at a fixed depth, runtimes are very consistent between executions; we consider it desirable to preserve this property. Langer et al. (2013) look at parallel branch and bound for integer programming, and explain that “reducing idle time by eagerly exploring as much of the tree as possible might be counter-productive by using compute resources for exploring sub-trees that might have been easily pruned later”. They evaluate two designs for work allocation, and observe that “even though Design A has better repeatability, the worst performance using Design B is better than the best performance using Design A”, and conclude that “Design B is the design of choice”. We do not wish to have to make this kind of trade-off.

4.1 A Low-Overhead, High-Diversity Work Splitting Mechanism

We must ensure early diversity (that is, weak heuristic commitment at the top of search), but retain a way of rebalancing subproblems. Our previous paper described (without detailed justification) a work donation mechanism, whereby subproblems are requeued when threads become idle; this is the light green dashed line in our graphs. This mechanism worked very well on the 12 core Intel machine used in

previous experiments, but it is clear that the approach is unsuitable for a 64 core AMD machine. We now introduce an alternative work splitting mechanism which is more suitable for larger numbers of cores.

Initially we split work at distance 1, as before, and place items in order onto a queue. When the queue is empty, and a thread first becomes idle, this thread then steals and requeues unstarted work from every other thread, splitting at a distance of 2. Finally, when the queue again becomes empty and a thread becomes idle, work is stolen and requeued with splitting at a distance of 3. This gives us all the search order benefits of distance 1 splitting, and the balance benefits of splitting at distance 3. In addition, if w is the number of workers and $|V|$ the number of vertices in the graph, it limits the queue size to be at worst $w \times |V|$ rather than $|V|^3$, and avoids enqueueing lots of eliminable nodes early on.

This scheme may be implemented with very little overhead, as follows. Each thread publishes three integers, describing its current position in search at the first three levels, and three corresponding flags. When work stealing takes place, the appropriate flag is set; when returning from a recursive call at the first three levels, the flag is checked to see whether the remainder of the work has already been stolen and requeued. In particular, this approach introduces no overhead below depth 3 (where most of the search time is spent), and is very unlikely to have any contended critical sections.

Results for this approach are shown using the dark green dashed line in the graphs—we see that in each case it is at least nearly as good as whichever other mechanism is best, and in many cases it is slightly better than the best alternative. (Since speedups are being given over a good sequential implementation, not over a parallel implementation run with one thread, it is legitimate to compare lines directly.) For all of the “brock400” graphs, “san400.0.9_1” and “gen400_p0.9_75”, we get the same superlinear speedup jumps as when splitting at distance 1. But “brock400_1”, “MANN_a45” and “C250.9” show that this approach is successful in addressing imbalances. With “p_hat700-3”, we initially behave similarly to splitting at distance 1, but as the number of threads increases and resplitting starts to have an effect earlier in search, our behaviour approaches that of distance 2 splitting. The graphs also clearly show that this technique scales well to 64 cores, and does not introduce noticeable overheads compared to static splitting mechanisms.

These experiments also suggest that with 64 cores, splitting at distance 3 is sufficient, and there is no benefit of continuing this to an arbitrary distance. However, this is not an intrinsic limitation, and a larger maximum depth can be used if necessary for larger numbers of cores or if unusually-shaped problem instances are found. Régim et al. (2013) suggest that the number of subproblems created only needs to grow linearly with the number of cores—increasing the effective splitting depth by resplitting gives a polynomial increase for linear cost.

4.2 Comparison to Off-the-Shelf Work Stealing

We now compare our new approach with an off-the-shelf randomised work stealing system, using the GCC implementation of Intel’s Cilk Plus extensions to C++. In Table 2 we present sequential and parallel runtimes for the larger solvable DIMACS instances, and the smallest group of BHOSLIB instances (these are selected so that parallel runtimes are at least half a second—below this point, measurements have considerable noise). Each parallel measurement is the average of ten runs: we show both runtimes, and the number of recursive calls made. For parallel measurements, we also give the standard deviation and the range of values observed, both expressed as a proportion.

For thirteen of the instances, resplitting both does less work and takes less time on average; in a further six cases, resplitting takes less time, but does more work. Cilk Plus does both less work and takes less time in six cases, and takes less time but does more work in one case. Additionally, when Cilk Plus is faster, it is never by more than 36%; in one case, resplitting wins by a factor of ten. (These results suggest that the assumption that each recursive call has the same cost is not entirely safe—this is not surprising, since we should not expect all colourings to be equally time-consuming. Additionally, our resplitting implementation counts requeued nodes multiple times, whereas the Cilk Plus implementation does not, so resplitting will show a slightly higher recursive call count when exactly the same amount of work is done.)

It may be possible to reduce the overheads incurred with Cilk Plus, through careful tuning. However, the difference between the average runtimes is not the most striking part of the results. When looking at the standard deviations and ranges, we see a more interesting picture: for many instances, the runtimes for Cilk Plus vary substantially between repeat runs, whereas the resplitting runtimes are much more consistent. (This is why we have not included Cilk Plus runtimes on our speedup graphs: the speedups

Table 2: A comparison of sequential and 64 core parallel runtimes for larger problem instances, using either resplitting or Cilk Plus for work stealing. On the first row we show runtimes, and on the second, search nodes. Parallel figures are an average of ten samples. The error is the standard deviation, and the figures in parentheses are the range, both expressed as a proportion. A \star indicates the best average.

Instance	Base	Resplitting	Cilk Plus
brock400_1	370.8 s 1.98×10^8	4.6 s ± 0.01 (0.99–1.02) \star $1.33 \times 10^8 \pm 0.00$ (1.00–1.00) \star	5.8 s ± 0.03 (0.96–1.04) $1.58 \times 10^8 \pm 0.04$ (0.95–1.05)
brock400_2	268.2 s 1.46×10^8	4.0 s ± 0.00 (0.99–1.01) $1.19 \times 10^8 \pm 0.01$ (0.99–1.01)	3.5 s ± 0.05 (0.92–1.06) \star $9.05 \times 10^7 \pm 0.05$ (0.92–1.07) \star
brock400_3	213.1 s 1.20×10^8	2.3 s ± 0.01 (0.99–1.01) \star $6.87 \times 10^7 \pm 0.01$ (0.99–1.01)	2.5 s ± 0.11 (0.87–1.22) $6.55 \times 10^7 \pm 0.12$ (0.87–1.25) \star
brock400_4	103.2 s 5.44×10^7	1.1 s ± 0.07 (0.95–1.19) $2.80 \times 10^7 \pm 0.05$ (0.93–1.13)	795 ms ± 0.14 (0.81–1.27) \star $1.34 \times 10^7 \pm 0.23$ (0.75–1.43) \star
brock800_1	6901.1 s 2.23×10^9	116.7 s ± 0.00 (1.00–1.00) \star $1.95 \times 10^9 \pm 0.00$ (1.00–1.00)	122.7 s ± 0.01 (0.98–1.02) $1.92 \times 10^9 \pm 0.02$ (0.98–1.03) \star
brock800_2	6890.8 s 2.24×10^9	102.8 s ± 0.00 (1.00–1.00) \star $1.64 \times 10^9 \pm 0.00$ (1.00–1.00) \star	117.3 s ± 0.01 (0.98–1.02) $1.80 \times 10^9 \pm 0.02$ (0.98–1.03)
brock800_3	6369.0 s 2.15×10^9	52.4 s ± 0.00 (1.00–1.01) \star $7.58 \times 10^8 \pm 0.00$ (1.00–1.01) \star	83.8 s ± 0.06 (0.94–1.13) $1.30 \times 10^9 \pm 0.07$ (0.92–1.16)
brock800_4	2413.2 s 6.40×10^8	48.6 s ± 0.00 (1.00–1.00) \star $7.15 \times 10^8 \pm 0.00$ (1.00–1.00) \star	56.0 s ± 0.06 (0.88–1.11) $8.03 \times 10^8 \pm 0.08$ (0.85–1.15)
C250.9	2184.6 s 1.08×10^9	41.5 s ± 0.04 (0.95–1.08) \star $1.17 \times 10^9 \pm 0.05$ (0.95–1.09) \star	43.8 s ± 0.07 (0.89–1.08) $1.17 \times 10^9 \pm 0.07$ (0.88–1.09)
C2000.5	33.9 h 1.82×10^{10}	2285.7 s ± 0.00 (1.00–1.01) $1.82 \times 10^{10} \pm 0.00$ (1.00–1.00) \star	2228.8 s ± 0.00 (1.00–1.00) \star $1.82 \times 10^{10} \pm 0.00$ (1.00–1.00)
DSJC1000_5	181.3 s 7.70×10^7	3.6 s ± 0.00 (0.99–1.01) \star $7.70 \times 10^7 \pm 0.00$ (1.00–1.00) \star	4.0 s ± 0.02 (0.98–1.04) $7.83 \times 10^7 \pm 0.00$ (1.00–1.00)
gen400_p0.9_65	156.0 h 1.76×10^{11}	4452.2 s ± 0.00 (1.00–1.00) \star $7.16 \times 10^{10} \pm 0.00$ (1.00–1.00) \star	6050.4 s ± 0.41 (0.57–2.04) $1.01 \times 10^{11} \pm 0.44$ (0.57–2.12)
gen400_p0.9_75	88.4 h 1.05×10^{11}	348.4 s ± 0.00 (1.00–1.01) \star $5.51 \times 10^9 \pm 0.00$ (0.99–1.01) \star	3532.5 s ± 0.35 (0.47–1.40) $6.01 \times 10^{10} \pm 0.35$ (0.49–1.42)
keller5	60.4 h 5.07×10^{10}	4022.3 s ± 0.00 (1.00–1.00) \star $5.07 \times 10^{10} \pm 0.00$ (1.00–1.00) \star	4232.1 s ± 0.00 (1.00–1.00) $5.07 \times 10^{10} \pm 0.00$ (1.00–1.00)
p_hat500-3	186.2 s 3.93×10^7	2.8 s ± 0.01 (0.98–1.01) \star $3.30 \times 10^7 \pm 0.01$ (0.99–1.01) \star	3.7 s ± 0.04 (0.95–1.09) $4.26 \times 10^7 \pm 0.04$ (0.96–1.08)
p_hat700-3	2299.2 s 2.82×10^8	49.2 s ± 0.00 (1.00–1.00) \star $3.33 \times 10^8 \pm 0.00$ (1.00–1.00)	51.5 s ± 0.02 (0.96–1.03) $3.33 \times 10^8 \pm 0.02$ (0.96–1.03) \star
p_hat1000-2	223.9 s 3.45×10^7	5.1 s ± 0.01 (0.98–1.01) $4.10 \times 10^7 \pm 0.00$ (1.00–1.01)	4.8 s ± 0.02 (0.97–1.03) \star $3.69 \times 10^7 \pm 0.02$ (0.98–1.04) \star
p_hat1000-3	323.8 h 1.30×10^{11}	5.7 h ± 0.00 (1.00–1.00) \star $1.28 \times 10^{11} \pm 0.00$ (1.00–1.00) \star	6.3 h ± 0.01 (0.98–1.01) $1.37 \times 10^{11} \pm 0.01$ (0.98–1.02)
p_hat1500-2	14.0 h 2.01×10^9	1144.2 s ± 0.00 (1.00–1.00) $2.54 \times 10^9 \pm 0.00$ (1.00–1.00)	972.7 s ± 0.02 (0.97–1.02) \star $2.24 \times 10^9 \pm 0.02$ (0.97–1.03) \star
sanr200_0.9	27.5 s 1.49×10^7	548 ms ± 0.04 (0.92–1.05) \star $1.43 \times 10^7 \pm 0.04$ (0.92–1.05) \star	865 ms ± 0.09 (0.90–1.11) $1.93 \times 10^7 \pm 0.11$ (0.84–1.18)
sanr400_0.7	96.1 s 6.44×10^7	1.8 s ± 0.01 (0.99–1.01) \star $6.54 \times 10^7 \pm 0.00$ (1.00–1.00)	2.1 s ± 0.02 (0.98–1.06) $6.47 \times 10^7 \pm 0.01$ (0.99–1.01) \star
frb30-15-1	999.6 s 2.92×10^8	15.9 s ± 0.00 (1.00–1.00) \star $2.54 \times 10^8 \pm 0.00$ (1.00–1.00)	16.5 s ± 0.02 (0.97–1.04) $2.45 \times 10^8 \pm 0.02$ (0.96–1.05) \star
frb30-15-2	1834.1 s 5.57×10^8	32.3 s ± 0.00 (1.00–1.01) $5.48 \times 10^8 \pm 0.00$ (1.00–1.00)	31.6 s ± 0.05 (0.93–1.07) \star $5.08 \times 10^8 \pm 0.06$ (0.92–1.08) \star
frb30-15-3	563.5 s 1.67×10^8	7.9 s ± 0.01 (0.99–1.01) \star $1.27 \times 10^8 \pm 0.00$ (1.00–1.01)	8.5 s ± 0.03 (0.95–1.03) $1.27 \times 10^8 \pm 0.03$ (0.95–1.05) \star
frb30-15-4	3136.8 s 9.91×10^8	38.7 s ± 0.00 (1.00–1.00) \star $6.79 \times 10^8 \pm 0.00$ (1.00–1.00) \star	58.4 s ± 0.01 (0.98–1.02) $9.92 \times 10^8 \pm 0.01$ (0.98–1.02)
frb30-15-5	913.3 s 2.83×10^8	21.9 s ± 0.04 (0.92–1.06) $3.80 \times 10^8 \pm 0.05$ (0.91–1.07)	16.9 s ± 0.05 (0.93–1.09) \star $2.72 \times 10^8 \pm 0.06$ (0.92–1.10) \star

are too chaotic to give meaningful data.) This is particularly visible with the “brock” and “gen” instances, which we have seen are very sensitive to the time taken to find the solution: for “gen400_p0.9_65”, our shortest observed runtime with Cilk Plus was 3477 s, and our longest was 12 315 s. By contrast, with resplitting, the shortest runtime was 4444 s and the longest was 4463 s. We do not consider the variability seen with Cilk Plus to be ideal behaviour.

Despite the lack of explicit diversity, we *do* often see superlinear speedups with Cilk Plus. We can explain this. The GCC Cilk Plus implementation tends to steal “earliest created” jobs first: behind the scenes, each worker has a deque, and jobs are enqueued and processed by the worker at one end (LIFO), but are stolen from the other (FIFO); victims for stealing are selected randomly. The rationale is that this is less likely to introduce contention, and that earlier-created jobs are likely to be larger (McCool et al., 2012). However, this has another benefit here: it introduces at least some diversity early in the search. Assuming the (1) subtree is non-trivial, the (2) subtree will always be stolen immediately, and a third thread will then either steal (1, 2) or (3), nondeterministically. This is an implementation detail, and not a specified behaviour, so it could change with a new release of the compiler—this could have disastrous effects for parallel branch and bound.

This accidental diversity also usually requires a higher number of threads before the incumbent can be found quickly. With 16 threads as opposed to 64, Cilk Plus is slower for every instance except “p-hat1500-2” (where it wins by less than 1 %), due to it requiring more work to find the solution.

4.3 Other Approaches

We have seen how this approach compares to that of Depolli et al. (2013): we improve the balance, whilst retaining the diversity. Compared to an off-the-shelf work stealing implementation, we often get better runtimes, and our performance is much more consistent between runs.

Compared to Xiang et al. (2013), we do not need to estimate up-front how large subtrees might be to obtain balance, and we do not need to do any calibration. Unfortunately Xiang et al. only considered three of the DIMACS graphs, all from the same family. None of these graphs have many avoidable nodes, and so the effects we discuss here are not seen in their results—this explains their consistently linear speedups (which are presented over a parallel algorithm, not a sequential one). Their approach also does not preserve any sequential order, and only shares updated incumbents when starting a new subproblem—when there are many avoidable nodes, this can cause a slowdown, both in theory (Trienekens, 1990) and in practice (McCreesh and Prosser, 2012). It is thus unclear what would happen with their approach on other DIMACS graph families, where scalability is not the only concern.

Compared to Moisan et al.’s parallel discrepancy search (Moisan et al., 2013), we are emphasising early diversity, not total number of discrepancies—that is, we believe Harvey and Ginsberg’s second claim (that heuristics are weak early on) is important for this problem, not their first (that the total number of wrong turns made is low). Another form of parallel limited discrepancy search, in a constraint programming setting, is discussed by Michel et al. (2009). They report that superlinear speedups are common for some problems when comparing parallel limited discrepancy search to sequential depth first search. However, such an approach risks introducing a slowdown (which they discuss), and again they consider the number of discrepancies rather than where those discrepancies occur.

Compared to the work by Batsyn, Maslov et al., we are trying to obtain a strong incumbent earlier in the search by using multiple paths through a single search tree, rather than a two stage process. This means that all of our work done is potentially contributing to a proof of optimality. We also do not need to select a pre-determined arbitrary time to run the first stage (Batsyn, Maslov et al. ran the heuristic for different amounts of time for different problems based upon pre-existing knowledge of the probable size of a solution and the difficulty of the problem), so our approach does not require special tuning before it can be used on “unseen” graphs. Note however that these approaches are not mutually exclusive—for problems which are expected to be particularly difficult, it would be possible to sacrifice one thread during early stages of our search process to perform a search using a non-exact algorithm.

Another approach using multiple search trees is cooperating local search (Clearwater et al., 1991), which has been demonstrated in a maximum clique context by Pullan et al. (2011). Here the aim is to combine multiple non-exact algorithms in parallel in the hopes of gaining the strengths of each; no attempt is made to prove optimality. This technique has also been used in graph colouring, where Lewandowski and Condon (1996) observe that “in general, having the processors work in parallel yields better colorings faster than simply using multiple independent runs”.

This theme also shows up in the ManySAT parallel SAT solver (Hamadi et al., 2009), where parallelism is used to counter the sensitivity of SAT solvers to initial tuning parameters. Hamadi et al. remark that “the performance of parallel solvers is usually better on SAT problems than on UNSAT ones”. This is because multiple search trees may make it easier to find a solution, but they are less helpful in proving that there is no solution (although things are not this simple for SAT, where learned clauses are shared). In contrast, in cases where a good solution can be found quickly, our approach allows all of the work done by every thread to contribute to a proof of optimality—this is why we typically get at least near linear speedups. Some SAT solvers also make use of randomised restarts (Gomes et al., 1998), again with the aim of avoiding heavy commitment to any particular portion of the search space; the emphasis here is upon finding a solution to a satisfiable instance more quickly.

Finally, one could use these experiments to support the view that *sequential* maximum clique algorithms should use some form of early discrepancy search. We would not dispute this, although we note firstly that more complicated sequential search algorithms introduce overheads which cannot be offset during the proof of optimality stage of search, and secondly that either way, every modern processor has multiple cores and we should be making use of this (Sutter, 2005). In any case, doing this would not invalidate our approach—we would simply see some of our superlinear speedups become linear speedups over a faster sequential baseline.

5 Conclusion

We have considered some of the design choices available when parallelising a state-of-the-art branch and bound algorithm for the maximum clique problem. We have shown that the irregularity of subproblem sizes can cause workload balance problems—this should not be surprising. But we also saw that different parallel search orders will often produce substantially different speedups, to the extent that balance is usually *not* the deciding factor for performance.

This sheds new light on Lai and Sahni’s claim that “anomalous behaviour will be rarely witnessed in practice” (Lai and Sahni, 1984). Both the authors and Depolli et al. *did* commonly encounter superlinear speedups, but only because our parallel search orders encouraged this. We also saw superlinear speedups when using Cilk Plus, but only because of an unintended effect of an unspecified implementation detail. We used our understanding of the behaviour of heuristics to provide an explanation: parallelism was introducing diversity into the search, avoiding a strong commitment to weak early heuristic advice. We then showed how to preserve this diversity whilst improving load balancing.

More generally, we have demonstrated that obtaining a parallel algorithm which seems to behave well most of the time should only be the first step in the algorithm design process. For heuristics, Hooker (1995) advocates a scientific approach to algorithm evaluation, rather than simple performance comparisons:

“Based on one’s insight into an algorithm, for instance, one might expect good performance to depend on a certain characteristic. How to find out? Design a controlled experiment that checks how the presence or absence of this characteristic affects performance. Even better, build an explanatory mathematical model that captures the insight, as is done routinely in other empirical sciences, and deduce from it precise consequences that can be put to the test.”

Such an approach has been helpful here too. By taking measurements inside search, and looking in depth at individual results, we gained insight into how to improve a parallel algorithm. We explained *why* the authors and Depolli et al. sometimes saw superlinear speedups, and how to preserve these speedups when modifying the parallel algorithm. We also showed why in certain cases Depolli et al. had parallelism limits, and why Batsyn, Maslov et al. failed to obtain an improvement by priming the search with a stronger incumbent for some problems.

We intend to apply this approach to other problems. In some ways, maximum clique is an “ideal” problem for this approach, and other problems may not have the same properties which could complicate matters. We are aware of a potentially unfavourable interaction between this approach and unbroken symmetries which must be addressed before some other hard problems may be considered. It is also not obvious whether Harvey and Ginsberg’s second claim (that heuristics are worst early in search) will always hold—in cases where it does not, it may be necessary to use heuristic quality information at runtime to direct parallel search order.

Acknowledgements

The authors wish to thank Matjaž Depolli for his comments and for making his source code available, Jennifer Boyle and David Manlove for their comments, and Douglas MacFarlane and Joe Sventek for their help with the hardware.

References

- David A. Bader, William E. Hart, and Cynthia A. Phillips. 2005. Parallel Algorithm Design for Branch and Bound. In *Tutorials on Emerging Methodologies and Applications in Operations Research*, HJ. G (Ed.). International Series in Operations Research & Management Science, Vol. 76. Springer New York, New York, NY, USA, 5–1–5–44. DOI:http://dx.doi.org/10.1007/0-387-22827-6_5
- Mikhail Batsyn, Boris Goldengorin, Evgeny Maslov, and Panos M. Pardalos. 2014. Improvements to MCS algorithm for the maximum clique problem. *Journal of Combinatorial Optimization* 27, 2 (2014), 397–416. DOI:<http://dx.doi.org/10.1007/s10878-012-9592-6>
- David Bergman, Andre A. Cire, Ashish Sabharwal, Horst Samulowitz, Vijay Saraswat, and Willem-Jan van Hoeve. 2014. Parallel Combinatorial Optimization with Decision Diagrams. In *Integration of AI and OR Techniques in Constraint Programming*, Helmut Simonis (Ed.). Lecture Notes in Computer Science, Vol. 8451. Springer International Publishing, 351–367. DOI:http://dx.doi.org/10.1007/978-3-319-07046-9_25
- Immanuel M. Bomze, Marco Budinich, Panos M. Pardalos, and Marcello Pelillo. 1999. The maximum clique problem. *Handbook of Combinatorial Optimization (Supplement Volume A)* 4 (1999), 1–74. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.6221>
- Sergiy Butenko and Wilbert E. Wilhelm. 2006. Clique-detection models in computational biochemistry and genomics. *European Journal of Operational Research* 173, 1 (2006), 1–17. DOI:<http://dx.doi.org/10.1016/j.ejor.2005.05.026>
- Yves Caniou, Philippe Codognet, Daniel Diaz, and Salvador Abreu. 2011. Experiments in Parallel Constraint-Based Local Search. In *Evolutionary Computation in Combinatorial Optimization*, Peter Merz and Jin-Kao Hao (Eds.). Lecture Notes in Computer Science, Vol. 6622. Springer Berlin Heidelberg, 96–107. DOI:http://dx.doi.org/10.1007/978-3-642-20364-0_9
- Peter Cheeseman, Bob Kanefsky, and William M. Taylor. 1991. Where the really hard problems are. In *Proceedings of the 12th international joint conference on Artificial intelligence - Volume 1 (IJCAI'91)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 331–337. <http://dl.acm.org/citation.cfm?id=1631171.1631221>
- Jens Clausen. 1997. Parallel branch and bound—principles and personal experiences. In *Parallel Computing in Optimization*, Athanasios Migdalas, Panos M. Pardalos, and Sverre Story (Eds.). Kluwer Academic Publishers, Norwell, MA, USA, Chapter 7, 239–267.
- Scott H. Clearwater, Bernardo A. Huberman, and Tad Hogg. 1991. Cooperative Solution of Constraint Satisfaction Problems. *Science* 254, 5035 (1991), 1181–1183. DOI:<http://dx.doi.org/10.1126/science.254.5035.1181>
- Teodor Gabriel Crainic, Bertrand Le Cun, and Catherine Roucairol. 2006. *Parallel Branch-and-Bound Algorithms*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 1–28. DOI:<http://dx.doi.org/10.1002/9780470053928.ch1>
- A. de Bruin, G.A.P. Kindervater, and H.W.J.M. Trienekens. 1995. Asynchronous parallel branch and bound and anomalies. In *Parallel Algorithms for Irregularly Structured Problems*, Afonso Ferreira and Jos Rolim (Eds.). Lecture Notes in Computer Science, Vol. 980. Springer Berlin Heidelberg, Berlin, Heidelberg, 363–377. DOI:http://dx.doi.org/10.1007/3-540-60321-2_29

- Jennifer Debroni, John D. Eblen, Michael A. Langston, Wendy Myrvold, Peter Shor, and Dinesh Weerapurage. 2011. A Complete Resolution of the Keller Maximum Clique Problem. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '11)*. SIAM, 129–135. <http://dl.acm.org/citation.cfm?id=2133036.2133047>
- Matjaž Depolli, Janez Konc, Kati Rozman, Roman Trobec, and Dušanka Janežič. 2013. Exact Parallel Maximum Clique Algorithm for General and Protein Graphs. *Journal of Chemical Information and Modeling* 53, 9 (2013), 2217–2228. DOI:<http://dx.doi.org/10.1021/ci4002525>
- John D. Eblen, Charles A. Phillips, Gary L. Rogers, and Michael A. Langston. 2011. The Maximum Clique Enumeration Problem: Algorithms, Applications and Implementations. In *Bioinformatics Research and Applications*, Jianer Chen, Jianxin Wang, and Alexander Zelikovsky (Eds.). Lecture Notes in Computer Science, Vol. 6674. Springer Berlin Heidelberg, Berlin, Heidelberg, 306–319. DOI: http://dx.doi.org/10.1007/978-3-642-21260-4_30
- Matteo Fischetti, Michele Monaci, and Domenico Salvagnin. 2014. Self-splitting of Workload in Parallel Computation. In *Integration of AI and OR Techniques in Constraint Programming*, Helmut Simonis (Ed.). Lecture Notes in Computer Science, Vol. 8451. Springer International Publishing, 394–404. DOI: http://dx.doi.org/10.1007/978-3-319-07046-9_28
- Michael R. Garey and David S. Johnson. 1990. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA.
- Carla P. Gomes, Bart Selman, and Henry Kautz. 1998. Boosting Combinatorial Search Through Randomization. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence (AAAI '98/IAAI '98)*. American Association for Artificial Intelligence, Palo Alto, CA, USA, 431–437. <http://dl.acm.org/citation.cfm?id=295240.295710>
- Youssef Hamadi, Said Jabbour, and Lakhdar Sais. 2009. ManySAT: a parallel SAT solver. *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)* 6 (2009), 245–262.
- William D. Harvey and Matthew L. Ginsberg. 1995. Limited Discrepancy Search. In *IJCAI (1)*. Morgan Kaufmann, San Francisco, CA, USA, 607–615.
- J.N. Hooker. 1995. Testing heuristics: We have it all wrong. *Journal of Heuristics* 1, 1 (1995), 33–42. DOI:<http://dx.doi.org/10.1007/BF02430364>
- Richard E. Korf. 1996. Improved Limited Discrepancy Search.. In *AAAI/IAAI, Vol. 1*. AAAI Press, Palo Alto, CA, USA, 286–291.
- Ten-Hwang Lai and Sartaj Sahni. 1984. Anomalies in parallel branch-and-bound algorithms. *Commun. ACM* 27, 6 (1984), 594–602.
- A Langer, R. Venkataraman, U. Palekar, and L.V. Kale. 2013. Parallel branch-and-bound for two-stage stochastic integer optimization. In *High Performance Computing (HiPC), 2013 20th International Conference on*. 266–275. DOI:<http://dx.doi.org/10.1109/HiPC.2013.6799130>
- Rudi Leroy, Mohand Mezma, Nouredine Melab, and Daniel Tuytens. 2014. Work Stealing Strategies For Multi-Core Parallel Branch-and-Bound Algorithm Using Factorial Number System. In *Proceedings of Programming Models and Applications on Multicores and Manycores (PMAM'14)*. ACM, New York, NY, USA, Article 111, 9 pages. DOI:<http://dx.doi.org/10.1145/2560683.2560694>
- Gary Lewandowski and Anne Condon. 1996. Experiments with parallel graph coloring heuristics and applications of graph coloring. *DIMACS series in discrete mathematics and theoretical computer science* 26 (1996), 309–334.
- Chu-Min Li, Zhu Zhu, Felip Manyà, and Laurent Simon. 2011. Minimum Satisfiability and Its Applications. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume One (IJCAI'11)*. AAAI Press, Palo Alto, CA, USA, 605–610. DOI:<http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-108>

- Guo-Jie Li and B. Wah. 1986. Coping with Anomalies in Parallel Branch-and-Bound Algorithms. *Computers, IEEE Transactions on* C-35, 6 (1986), 568–573. DOI:<http://dx.doi.org/10.1109/TC.1986.5009434>
- Evgeny Maslov, Mikhail Batsyn, and Panos M. Pardalos. 2013. Speeding up MCS Algorithm for the Maximum Clique Problem with ILS Heuristic and Other Enhancements. In *Models, Algorithms, and Technologies for Network Analysis*, Boris I. Goldengorin, Valery A. Kalyagin, and Panos M. Pardalos (Eds.). Springer Proceedings in Mathematics & Statistics, Vol. 59. Springer New York, New York, NY, USA, 93–99. DOI:http://dx.doi.org/10.1007/978-1-4614-8588-9_7
- Michael McCool, James Reinders, and Arch Robison. 2012. *Structured parallel programming: patterns for efficient computation*. Elsevier.
- Ciaran McCreesh and Patrick Prosser. 2012. Distributing an Exact Algorithm for Maximum Clique: maximising the costup. *CoRR* abs/1209.4560 (2012).
- Ciaran McCreesh and Patrick Prosser. 2013a. Greedy Graph Colouring is a Misleading Heuristic. *CoRR* abs/1310.7741 (2013).
- Ciaran McCreesh and Patrick Prosser. 2013b. Multi-Threading a State-of-the-Art Maximum Clique Algorithm. *Algorithms* 6, 4 (2013), 618–635. DOI:<http://dx.doi.org/10.3390/a6040618>
- Laurent Michel, Andrew See, and Pascal Van Hentenryck. 2009. Transparent Parallelization of Constraint Programming. *INFORMS Journal on Computing* 21, 3 (2009), 363–382. DOI:<http://dx.doi.org/10.1287/ijoc.1080.0313>
- Thierry Moisan, Jonathan Gaudreault, and Claude-Guy Quimper. 2013. Parallel Discrepancy-Based Search. In *Principles and Practice of Constraint Programming*, Christian Schulte (Ed.). Lecture Notes in Computer Science, Vol. 8124. Springer Berlin Heidelberg, Berlin, Heidelberg, 30–46. DOI:http://dx.doi.org/10.1007/978-3-642-40627-0_6
- Panos M. Pardalos, Jonas Rappe, Mauricio, and Mauricio G.C. Resende. 1998. An Exact Parallel Algorithm For The Maximum Clique Problem. In *In High Performance and Software in Nonlinear Optimization*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 279–300.
- Michael Poldner and Herbert Kuchen. 2008. Algorithmic Skeletons for Branch and Bound. In *Software and Data Technologies*, Joaquim Filipe, Boris Shishkov, and Markus Helfert (Eds.). Communications in Computer and Information Science, Vol. 10. Springer Berlin Heidelberg, 204–219. DOI:http://dx.doi.org/10.1007/978-3-540-70621-2_17
- Patrick Prosser. 2012. Exact algorithms for maximum clique: a computational study. *Algorithms* 5, 4 (2012), 545–587. DOI:<http://dx.doi.org/10.3390/a5040545>
- Patrick Prosser and Chris Unsworth. 2011. Limited discrepancy search revisited. *Journal of Experimental Algorithmics (JEA)* 16 (2011), 1–6.
- Wayne Pullan, Franco Mascia, and Mauro Brunato. 2011. Cooperating local search for the maximum clique problem. *Journal of heuristics* 17, 2 (2011), 181–199.
- V.J. Rayward-Smith, S.A. Rush, and G.P. McKeown. 1993. Efficiency considerations in the implementation of parallel branch-and-bound. *Annals of Operations Research* 43, 2 (1993), 123–145. DOI:<http://dx.doi.org/10.1007/BF02024489>
- Jean-Charles Régin. 2003. Using Constraint Programming to Solve the Maximum Clique Problem. In *Principles and Practice of Constraint Programming CP 2003*, Francesca Rossi (Ed.). Lecture Notes in Computer Science, Vol. 2833. Springer Berlin Heidelberg, Berlin, Heidelberg, 634–648. DOI:http://dx.doi.org/10.1007/978-3-540-45193-8_43
- Jean-Charles Régin, Mohamed Rezgui, and Arnaud Malapert. 2013. Embarrassingly Parallel Search. In *Principles and Practice of Constraint Programming*, Christian Schulte (Ed.). Lecture Notes in Computer Science, Vol. 8124. Springer Berlin Heidelberg, Berlin, Heidelberg, 596–610. DOI:http://dx.doi.org/10.1007/978-3-642-40627-0_45

- Gergely Regula and Béla Lantos. 2013. Formation Control of Quadrotor Helicopters with Guaranteed Collision Avoidance via Safe Path. *Electrical Engineering and Computer Science* 56, 4 (2013), 113–124.
- Ryan A. Rossi, David F. Gleich, Assefaw Hadish Gebremedhin, and Md. Mostofa Ali Patwary. 2013. A Fast Parallel Maximum Clique Algorithm for Large Sparse Graphs and Temporal Strong Components. *CoRR* abs/1302.6256 (2013).
- Pablo San Segundo, Alvaro Lopez, and Mikhail Batsyn. 2014. Initial Sorting of Vertices in the Maximum Clique Problem Reviewed. In *Learning and Intelligent Optimization*, Panos M. Pardalos, Mauricio G.C. Resende, Chrysafis Vogiatzis, and Jose L. Walteros (Eds.). Springer International Publishing, 111–120. DOI:http://dx.doi.org/10.1007/978-3-319-09584-4_12
- Pablo San Segundo, Fernando Matia, Diego Rodriguez-Losada, and Miguel Hernando. 2013. An improved bit parallel exact maximum clique algorithm. *Optimization Letters* 7, 3 (2013), 467–479. DOI:<http://dx.doi.org/10.1007/s11590-011-0431-y>
- Pablo San Segundo, Diego Rodríguez-Losada, and Agustín Jiménez. 2011. An exact bit-parallel algorithm for the maximum clique problem. *Comput. Oper. Res.* 38, 2 (Feb. 2011), 571–581. DOI:<http://dx.doi.org/10.1016/j.cor.2010.07.019>
- Peter Sanders. 1995. Better algorithms for parallel backtracking. In *Parallel Algorithms for Irregularly Structured Problems*, Afonso Ferreira and Jos Rolim (Eds.). Lecture Notes in Computer Science, Vol. 980. Springer Berlin Heidelberg, Berlin, Heidelberg, 333–347. DOI:http://dx.doi.org/10.1007/3-540-60321-2_27
- Herb Sutter. 2005. The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software. *Dr. Dobbs's Journal* 30, 3 (2005).
- E. Tomita and T. Kameda. 2007. An efficient branch-and-bound algorithm for finding a maximum clique with computational experiments. *Journal of Global Optimization* 37, 1 (2007), 95–111.
- Etsuji Tomita and Tomokazu Seki. 2003. An efficient branch-and-bound algorithm for finding a maximum clique. In *Proceedings of the 4th international conference on Discrete mathematics and theoretical computer science (DMTCS'03)*. Springer-Verlag, Berlin, Heidelberg, 278–289. <http://dl.acm.org/citation.cfm?id=1783712.1783736>
- Etsuji Tomita, Yoichi Sutani, Takanori Higashi, Shinya Takahashi, and Mitsuo Wakatsuki. 2010. A Simple and Faster Branch-and-Bound Algorithm for Finding a Maximum Clique. In *WALCOM: Algorithms and Computation*, Md.Saidur Rahman and Satoshi Fujita (Eds.). Lecture Notes in Computer Science, Vol. 5942. Springer Berlin Heidelberg, Berlin, Heidelberg, 191–203. DOI:http://dx.doi.org/10.1007/978-3-642-11440-3_18
- Harry W.J.M. Trienekens. 1990. *Parallel Branch and Bound Algorithms*. Ph.D. Dissertation. Erasmus University Rotterdam.
- Toby Walsh. 1997. Depth-bounded Discrepancy Search. In *IJCAI*. Morgan Kaufmann, San Francisco, CA, USA, 1388–1395.
- Jingen Xiang, Cong Guo, and A. Aboulnga. 2013. Scalable maximum clique computation using MapReduce. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*. 74–85. DOI:<http://dx.doi.org/10.1109/ICDE.2013.6544815>